Limits of visual communication: the effect of signal-to-noise ratio on the intelligibility of American Sign Language

M. Pavel

Department of Psychology, Stanford University, Stanford, California 94305

George Sperling, Thomas Riedl, and August Vanderbeek

Human Information Processing Laboratory, Department of Psychology, New York University, New York, New York 10003

Received April 23, 1987; accepted August 19, 1987

To determine the limits of human observers' ability to identify visually presented American Sign Language (ASL), the contrast s and the amount of additive noise n in dynamic ASL images were varied independently. Contrast was tested over a 4:1 range; the rms signal-to-noise ratios (s/n) investigated were $s/n = 1/4, 1/2, 1$, and ∞ (which is used to designate the original, uncontaminated images). Fourteen deaf subjects were tested with an intelligibility test composed of 85 isolated ASL signs, each 2–3 sec in length. For these ASL signs (64×96 pixels, 30 frames/sec), subjects' performance asymptotes between $s/n = 0.5$ and 1.0; further increases in s/n do not improve intelligibility. Intelligibility was found to depend only on s/n and not on contrast. A formulation in terms of logistic functions was proposed to derive intelligibility of ASL signs from s/n, sign familiarity, and sign difficulty. Familiarity (ignorance) is represented by additive signal-correlated noise; it represents the likelihood of a subject's knowing a particular ASL sign, and it adds to s/n. Difficulty is represented by a multiplicative difficulty coefficient; it represents the perceptual vulnerability of an ASL sign to noise and it adds to $\log(s/n)$.

INTRODUCTION

How good must the quality of visual information be in order for images to be useful? Visual scientists would like to know the answer to this question in order to describe the human visual system better. Engineers and image-processing-systems designers would like to know so that they may build the most effective systems. Clearly, this question has both practical and theoretical importance.

From the practical point of view, the popularity of movies, television, and video games attests to the fact that people like to watch dynamic, electronically or optically transformed two-dimensional images, sometimes even in preference to natural scenes! Currently, most of the effort devoted to the technology of video and image processing is directed at producing images of the highest possible quality, requiring high information transmission rates. The objective is high-fidelity reproduction of original images.

Aside from aesthetic considerations, high image quality may not be necessary for successful performance on many tasks such as visual inspection, motor control, and communication. An indication of how low the requirements may be for a successful execution of many tasks is suggested by a number of recent studies, including those of the capabilities of visually impaired patients.[1–4] These findings are supported by several studies of the abilities of normals to interpret static low-bandwidth images[2,3,5–8] and by research on American Sign Language (ASL) communication with low bandwidth.[9–16]

From a theoretical point of view, researchers who intend to characterize and model the visual perceptual system are interested in what information in images is used by the visual system for different tasks. The simplest theory of perceptual communication is based on the assumption that the information is extracted from images by an optimal receiver using feature detectors similar to templates (matched filters). The performance ought to be characterizable within the classical information-theoretic framework (e.g., see Refs. 17 and 18). In particular, the amount of information should be a function of the signal-to-noise ratio.

In the present study we examine the amount of information necessary for visual communication with American Sign Language (ASL) used by the deaf community. A significant proportion of the deaf community uses manual communication forms such as ASL. ASL is a complex language in which meaning is conveyed by gestures and motions that are concatenated under syntactic constraints. For the purpose of this work it is important to note that the potential of communicating with ASL is similar to that of speech. For example, as a language in use for everyday communication, ASL is used by the deaf community at the same rate of communication as a spoken language.[19]

We assume that the amount of useful information can be inferred from the level of performance on ASL communication tasks. Given this assumption, there are at least two ways to determine the limits of necessary image quality. One way involves reducing the amount of information in the original image by filtering or removing information. The studies referred to above all use this method. The second approach involves adding noise (i.e., random signals) that mask image information. Although this approach has been widely used in audition,[20] noise has been used with dynamic ASL images only in a study of location discrimination.[21] In the present study, we use additive noise to reduce the amount of information in images of ASL signs in order to

0740-3232/87/122355-11$02.00

assess whether the signal-to-noise ratio describes the performance of ASL signers (the signal-to-noise hypothesis) or whether the noise and the signal combine in some more complicated way to determine ASL intelligibility.

First, we show how the signal-to-noise ratio (s/n) applies to theoretical descriptions of receiver performance and perception. The performance of a communication system is characterized by the probability of correctly identifying messages as a function p of the signal rms power s and the noise rms power n. For a large class of realizable optimal receivers (e.g., matched filters), the degrading effect of additive noise on receiver performance can be expressed in terms of a monotonic function f of the ratio of the signal power to the noise power:

$$\text{Prob}\{\text{correct detection}\} = p(s, n) = f(s/n) \qquad (1)$$

for noise with a constant spectrum. The consequence of Eq. (1) is that increasing both the noise and the signal power by the same factor will not change the probability of correct detection. We note in passing that the signal-to-noise invariance is also consistent with the classical information-theoretic analysis of transmission systems in which the capacity of a channel is a function of the signal-to-noise ratio.[22]

Perceptual processes may also be viewed and studied as communication systems. Empirical findings in psychological research indicate that the addition of noise to stimuli frequently results in signal-to-noise invariance. Such invariances enable us to model perceptual processes as optimal receivers.[23] In human signal detection, performance also is characterized by the probability of obtaining a correct response. The signal-to-noise invariance is expressed as a constant ratio

$$s/n = k_p, \qquad (2)$$

where k_p is a constant that depends on p but is independent of s and n. When Eq. (2) holds for all values of p, then the probability of detection is a function of the ratio s/n. With respect to the signal-to-noise invariance expressed in Eq. (1), taking the inverse of both sides of Eq. (1) gives $k_p = f^{-1}(p)$, where f is the monotonic function from Eq. (1).

Similar invariance is observed in experiments in which subjects are asked to detect incremental intensity signals ΔI on backgrounds I. In these experiments, the probability of detection p is said to follow Weber's law[24,25] whenever the following holds: $\Delta I/I = k_p$. Alternatively, Weber's law may be regarded as a case in which the additive background noise is simply proportional to the background intensity.

In psychoacoustics, the signal-to-noise ratio is a good description of the human ability to detect auditory signals (e.g., pure tones) in noise. Signal-to-noise invariance holds also in visual masking tasks in which human observers detect the presence of simple visual patterns on a noisy background.[26] One goal of this work is to determine whether signal-to-noise invariance holds for complex visual tasks with dynamic stimuli.

We proceed as follows. First, we describe the development of an intelligibility test and its use to measure intelligibility of ASL for various combinations of signal and noise levels. The results of this experiment are used to test the signal-to-noise hypothesis and to describe the empirical relationship between the signal-to-noise ratio and intelligibility by an explicit function. In the final section a model is derived, based on the empirical results, to predict intelligibility from the signal-to-noise ratio, the difficulty of individual ASL signs, and the subject's skill.

Speech intelligibility refers to the probability of correctly identifying spoken messages, usually isolated words. The stimuli used to measure speech intelligibility have been carefully chosen and standardized so that this measure will predict the performance of audio communication systems. Similarly, it would be useful to develop an intelligibility measure for ASL that is defined as the probability of correctly identifying isolated ASL signs. This intelligibility measure could then be used to predict the performance of ASL communication systems.

Speech intelligibility has been found to depend on the signal-to-noise ratio. For example, Hawkins and Stevens[20] found that an increase in the power of background noise must be accompanied by a proportional increase in speech power in order to maintain constant intelligibility. Consequently, the effectiveness of audio communication systems can be described by their signal-to-noise ratios.

To determine whether the signal-to-noise invariance holds for the ASL intelligibility, we must first define signal and noise power for images and video signals. We must then develop an intelligibility measure for ASL. With these tools, we can test the hypothesis that the intelligibility of ASL degraded by additive noise is a function of the signal-to-noise ratio. In practice, signal-to-noise invariance enables us to predict intelligibility of ASL when it is degraded by noise. Theoretically, the relationship between ASL intelligibility and noise presents a useful way of describing the information in the images and consequently of characterizing, in informational terms, the process of perceiving ASL.

METHOD

In this section we describe the construction of the intelligibility test, digital processing and stimulus generation, and the procedure for the experimental testing of ASL intelligibility.

Apparatus

The recording and testing apparatus is outlined in Fig. 1. A video camera (JVC s100u) was used to record sign sequences monochromatically on a video cassette recorder (Sony Beta Max SLO 323, Beta I format). Images were digitized by transferring them first to a video motion analyzer (Sony SVM 1010), which permitted individual frames to be accessed. An image processor (Grinnell GMR 27) then digitized the image frames to a spatial resolution of 512 × 512 with 8 bits of (nominal) luminance resolution. Frames were uploaded to a digital computer (VAX 11/750) and subsequently processed by using the HIPS system[27,28] running under the UNIX 4.1bsd operating system. The individual images (frames) were reduced in x and y pixel dimensions by a factor of 4 to produce the source frames from which all stimuli were derived.

After processing, the images were downloaded from the computer to the video cassette recorder (VCR). The VCR was used to drive the monitor (Koyo 9 in. TMC-9M with standard NTSC B&W raster) that displayed the signs to the subjects (see Fig. 2). Analog contrast and brightness were set to minimize nonlinearity, saturation, and cutoff; settings

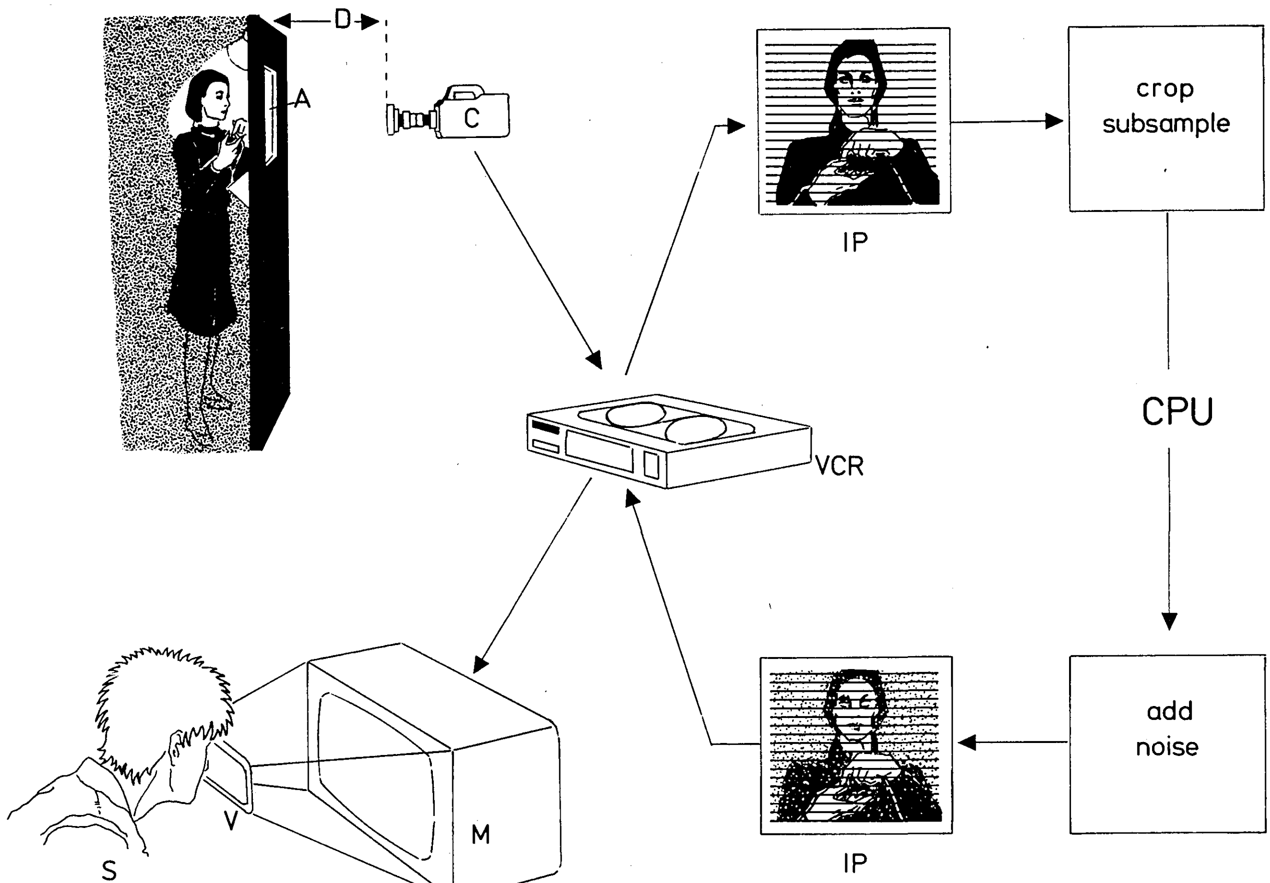

Fig. 1. The production of test stimuli. The signer is illuminated by optimally located lamps and photographed through a 12 in. × 18 in. (30.5 cm × 45.7 cm) aperture (A) by a video camera (C) (located at D = 4.44 m) and recorded on a VCR. The output of the VCR is converted into a machine-readable format by a Grinnell image processor (IP). A DEC VAX 11/750 computer is used to crop, subsample, and add noise to the image, which then is reconverted into video format by the IP, rerecorded on the VCR, and viewed by subjects on location on a monitor (M) through a viewing hood (V).

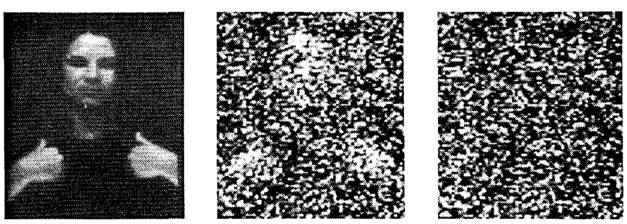

Fig. 2. Single frames taken from the ASL sequence "animal" with added noise. (a) $s/n = \infty$; (b) $s/n = 1$; (c) $s/n = 0$ (pure noise). Because the signal is correlated between frames and the noise is not, the single-frame static illustration (b) appears to be of a lower image quality than the dynamic images viewed by the subjects.

were checked daily and maintained with a photometer. A viewing hood was attached to the monitor to minimize screen reflection and ambient light and to control the viewing distance. Subjects viewed the images at a distance of 56 cm from the video screen, resulting in a visual angle of the stimulus field of approximately 4.4 deg horizontally and 6.6 deg vertically.

The processing of the stimuli consisted of digital preprocessing followed by the generation of sequences of test stimuli with specified contrasts and added noise. The original (512 × 512 pixel) images were cropped and reduced by block averaging to produce images measuring 96 × 64 pixels. These new individual frames were then concatenated to be presented as 30-frame/sec sequences 2, 2.5, and 3 sec long because the particular signs have different durations. The actual frame was reenlarged by a factor of 2 and consisted of 192 scan lines produced by two interlaced fields.

The first step in generating the stimulus sequences was to determine the actual rms contrast of each sequence. The actual rms contrast $\hat{C}_k$ of the kth frame is defined as the rms deviation of the pixel luminance from the mean pixel luminance of that frame, divided by the frame's mean luminance:

$$\hat{C}_k = \left[\frac{\sum_{i=1}^{m} (v_k(i) - \bar{v}_k)^2}{(m-1)\bar{v}_k^{\,2}} \right]^{1/2}, \tag{3}$$

where m is the total number of pixels in each frame (m = rows × columns), $v_k(i)$ is the luminance value of pixel i in frame k ($k = 1, \ldots, K$), and $\bar{v}_k = \sum_{i=1}^{m} v_k(i)/m$ is the average pixel intensity in the kth frame. The average actual rms contrast $\hat{C}$ of a sequence of K frames is the average of $\hat{C}_k$ over that sequence:

$$\hat{C} = \frac{1}{K} \sum_{k=1}^{K} \hat{C}_k.$$

Because the actual contrast of each ASL sequence is slightly different, we will use the term contrast to designate the normalized contrast C, where C is defined as unity for the original sequence. Thus the luminance of pixels in an ASL sequence with contrast C is determined from the original ASL sequence by

$$v_k(i) = \bar{v}_k + C[v_k(i) - \bar{v}_k].$$

The additive noise used in this experiment was distributed normally (Gaussian noise) with a uniform power spectrum over the entire range of spatial and temporal frequencies, limited only by the corresponding sampling rates; 14 cycles/deg of visual angle and 15 Hz. The noise rms values N were determined from the actual rms contrast $\hat{C}$ of the sequence and the desired signal-to-noise ratio s/n:

$$\hat{N} = \frac{C\hat{C}}{s/n}. \tag{4}$$

The noise was generated by sampling from a normal distribution with a mean of zero and a standard deviation $\hat{N}$. This noise was then algebraically added, pixel by pixel, to the intensity values of the pixels of the original image.

Our image-processing system has an internal intensity range of 0–255. The ranges of contrast and s/n used were chosen to maximize the probability that the signal with noise remained within the allowable range 0–255 of pixel values while utilizing the largest possible dynamic range. This was of concern because saturation and cutoff (values outside the range 0–255) produce distortion that is difficult to characterize in terms of additive noise. None of the sign sequences contained more than 3% saturated pixels.

Intelligibility Test Development

The intelligibility test was designed to measure the ability of ASL signers to identify individual signs isolated from context. Identifying isolated signs is a stringent test of performance because in natural ASL conversation, signers depend greatly on context. What follows is a brief description of the intelligibility test development process.

1. *Stimulus Selection*

Initially, 112 candidate signs were selected from a previously constructed list of 300 ASL signs that were designed to represent various classes of words as well as a range of hand shapes, locations, and movements. An attempt was made to avoid unusual or ambiguous signs.

2. *Preselection of 88 Test Signs*

In a small preselection experiment, two expert observers viewed each sign, without noise and with normal contrast. They were asked to indicate in writing what sign had been presented. Signs to which both subjects responded incorrectly were eliminated from the test.

3. *Difficulty Rating*

In order to produce groups of signs of approximately equal difficulty and noise susceptibility, a preliminary intelligibility test was run. The stimuli in this preliminary test were the images of the preselected ASL signs. Digitally generated Gaussian noise was added to the signs (on a per-pixel basis) such that the s/n was 1/2, and the original contrast was unchanged ($C = 1$). Two subjects were asked to identify each of the signs. Each stimulus was then assigned an ordinal difficulty score according to the performance of the two subjects as follows: (1) both subjects correct (easy), (2) one subject correct (difficult), and (3) both subjects incorrect (most difficult). The difficulty scores were used to divide the stimuli into 11 blocks of eight signs in order to equalize the difficulty of each block. Three other factors were balanced (as much as possible) between blocks: symmetry (i.e., are the left- and right-hand shapes and movements the same?), location (where in space is the sign produced?), and hand shape.

4. *Final Sign Set*

The resulting test stimuli consisted of 88 common ASL signs (one sign was repeated). There were 43 nouns, 30 verbs, 12 adjectives, 1 adverb and 1 conjunction. The complete set is included in Table 1. Typical of the nouns, verbs, and adjec-

Table 1. Representation of the Difficulty of ASL Signs

Stimulus (ASL)	Preliminary Difficulty	Word Frequency	Probability Correct	No. of Trials	Difficulty ($\log_e$)
Accident	1	33	92.86	14	0.552
Animal	2	68	82.14	28	−0.380
Apple	3	9	78.57	14	0.141
Bear	2	57	100.00	13	<−0.435
Because	1	883	100.00	14	<0.098
Behind	1	258	7.14	14	0.311
Believe	1	200	100.00	14	<−0.397
Boring	1	5	0.00	14	>0.765
Boss	1	5	85.71	14	0.031
Bread	1	41	23.08	13	0.007
Bug	2	4	50.00	14	1.125
Challenge	1	36	85.71	14	0.724
Cheese	3	9	92.86	14	−0.142
Color	1	141	100.00	13	<−0.435
Cop	1	15	100.00	14	<0.098
Country	3	324	92.86	14	0.552
Criticize	1	4	50.00	14	−0.262
Daily	2	122	42.86	14	0.496
Deaf	1	12	100.00	14	<−0.595
Disbelieve	2	3	92.31	13	0.571
Don't-want	1	489	78.57	14	−0.552

Table 1. Continued

Stimulus (ASL)	Preliminary Difficulty	Word Frequency	Probability Correct	No. of Trials	Difficulty ($\log_e$)
Emphasize	1	20	0.00	14	>0.765
Eye	1	122	100.00	14	<−0.595
Finish	1	39	100.00	14	<0.098
Flag	2	16	85.71	14	0.724
Flower	1	23	35.71	14	−0.130
Follow	2	97	71.43	14	0.920
Football	2	36	78.57	14	0.141
Friday	1	60	14.29	14	0.139
Girl	2	220	78.57	14	0.834
Good	3	807	0.00	14	>0.765
Guilty	1	29	0.00	14	>0.765
Home	2	547	100.00	14	<−0.595
Hospital	2	110	78.57	14	0.141
Improve	1	39	100.00	14	<0.098
Jump	2	24	85.71	14	0.031
Kill	3	63	7.14	14	0.311
Leave	3	205	35.71	14	−0.130
Letter	1	145	92.86	14	−0.516
Lousy	3	12	100.00	13	<0.099
Love	1	232	92.86	14	0.552
Machine	1	103	71.43	14	−0.466
Member	2	137	100.00	13	<0.099
Mother	1	216	100.00	14	<0.098
Noon	1	25	21.43	14	0.029
Our	1	1252	100.00	14	<0.098
Owe	1	10	85.71	14	0.724
Paper	1	157	100.00	14	<−0.595
Past/ago	2	271	85.71	14	0.031
Pay	2	172	7.14	14	0.311
Penny	3	25	92.86	14	−0.142
Plan	1	205	100.00	14	<0.098
Pour	2	9	64.29	14	0.345
Preach	1	8	7.14	14	0.311
Pregnant	3	8	92.31	13	0.828
Punish	2	3	92.86	14	0.552
Read	1	173	7.14	14	0.311
Red	1	197	61.54	13	1.021
Relax	1	19	57.14	14	0.367
School	1	492	78.57	14	0.141
Screwdriver	1	5	14.29	14	0.139
Sergeant	3	0	71.43	14	0.920
Secret	3	78	71.43	14	0.920
Sheep	2	23	14.29	14	0.832
Short	2	212	100.00	13	<−0.435
Sit	2	67	28.57	14	−0.057
Sorry	1	48	100.00	13	<−0.595
Start	1	154	100.00	14	<−0.595
Suffer	2	33	100.00	14	<0.098
Summer	2	134	78.57	14	0.834
Talk	1	154	7.14	14	0.311
Telegraph	2	21	42.86	14	0.496
Tempt	1	2	78.57	14	0.375
Think	3	433	92.86	14	−0.142
Tobacco	2	19	78.57	14	0.141
Tomato	2	9	35.71	14	0.563
Train	1	83	85.71	14	0.724
Tree	1	59	92.86	14	0.552
Ugly	1	21	28.57	14	−0.057
Uncle	1	57	100.00	14	<0.098
Understand	1	137	92.86	14	−0.142
Week	1	275	100.00	14	<0.098
Wife	3	228	100.00	14	<0.098
Wolf	3	6	0.00	14	>0.765
World	1	787	92.31	13	−0.124
Wrestling	1	1	50.00	14	−0.262
Yesterday	2	83	100.00	14	<0.098

tives used were, respectively, girl, cheese, and tomato; sit, talk, and tempt; and good, lousy, and ugly. A native ASL signer was videotaped signing the isolated ASL signs. The signs were performed without lip movement or other facial expression in order to provide an estimate of pure ASL communication (even though facial gestures usually accompany ASL). The same starting and final positions (arms folded) were used as context for each isolated sign.

Stimulus Parameter Determination

In order to determine the range of contrasts to be used for the study, a preliminary contrast-sensitivity test was run. The test signs were informally presented at a variety of contrasts C without added noise. Subjects performed at high levels even with C equal to only 1/16.

Stimulus Block Ordering

Subjects viewed a stimulus tape on which all the experimental stimuli were displayed in a predetermined order. Each stimulus tape consisted of nine blocks of the main 3×3 experimental conditions and two blocks of additional conditions (zero-noise control and an extra contrast condition). Each block consisted of eight signs. Each main block was recorded at one of three contrast levels and one of three signal-to-noise ratios, thus forming a 3×3 design. The main s/n conditions were produced by the following combinations of signal and noise (s, n) contrasts:

$$s/n = 1.00: \quad (0.5/0.5), (0.25/0.25), (0.125/0.125);$$

$$s/n = 0.50: \quad (0.5/1.0), (0.25/0.50), (0.125/0.250);$$

$$s/n = 0.25: \quad (0.5/2.0), (0.25/1.00), (0.125/0.500).$$

The signal and noise levels were chosen to avoid both large values of $s + n$, where saturation would occur, and low values of s, where intensity quantization and other extraneous sources of system noise would become significant. Of the two additional blocks, one displayed a pure signal at a signal contrast of 1 (original contrast). When no noise was added the actual $s/n \approx 64$, obtained by careful measurement,[29] was designated as $s/n = \infty$. This block provided a performance baseline. The other block was a fourth contrast condition for $s/n = 1$ with $C = 1$.

In order to counterbalance any possible systematic effect of block difficulty on performance, three different VCR tapes were created and used in the study. Each block of signs was produced with a different contrast level on each tape. Thus, at each s/n level, all comparisons of different contrast levels were completely balanced; that is, each contrast level was tested with precisely the same ASL signs as were the other two contrast levels, and every sign was seen at all three contrast levels (by different subjects).

At all contrast levels and s/n ratios, including $s/n = \infty$, the signal was quantized to 16 intensity levels so that there would be no confounding of the number of intensity levels with the range of signal contrast. The overall stimulus $s + n$ was quantized at 256 levels.

Subjects

The subjects were 14 deaf, fluent-signing adults who used ASL as their primary mode of communication. Four subjects were native signers who learned ASL as their first language, seven subjects first learned sign language before the age of 7 years, and three subjects learned sign language before the age of 20 years. Of the five males and nine females involved, seven were born deaf. The other seven subjects lost their hearing no later than the age of 3 years. The average age of the subjects was 60 years, with the ages ranging from 40 to 76 years. The subjects' occupations included those of housewife, architect, printer, proofreader, and ASL teacher.

Procedure

Subjects first answered a written questionnaire that inquired about their sign-language background. The subjects then viewed an instructional video tape. On this video tape, a native deaf signer explained the nature of the test materials and response requirements. The video tape also showed an example of visual noise. The brightness and contrast controls were set identically for all subjects. Subjects were instructed first to imitate the sign presented and then to write an English gloss for the perceived sign on the score sheet provided. Any questions the subjects had were answered by the experimenter, who was fluent in ASL. Sessions lasted approximately 40 min.

Scoring Responses

The subjects' written responses were scored for accuracy. The nature of ASL requires careful interpretation of the subjects' responses. In particular, certain signs can elicit different English word responses, analogous to synonyms in English. For example, when shown the sign "relax," a subject can correctly respond with any of various different written answers: "relax," "content," "satisfy," and "relieve." Each of these responses is equally acceptable. In other instances, one articulated sign can have two totally different meanings. As with homonyms in English, the context in which the sign is used clarifies its meaning. For example, one articulated sign can elicit the written responses "pepper" or "preach." Either of these answers would be scored as correct for that context-dependent sign. All synonym and homonym responses were scored as correct. Additionally, over the course of time, some signs have changed their meaning, while retaining their same form. For example, the sign articulation that presently means "bread" formerly meant "how." Both English responses, "bread" and "how," were scored as correct.

RESULTS AND DISCUSSION

Overall Performance

The subjects' performance is summarized in terms of the proportion of correct responses for each stimulus item and each contrast level and noise level. The overall proportion of incorrect identifications was 32% in the total of 1181 trials. A few of the trials were excluded from the analysis either because the observer did not see the stimulus altogether or because the response was uninterpretable. Of the incorrect trials, 58% were incorrect words (false alarms), and the remaining 42% were omissions. Thus our subjects were somewhat conservative; had they guessed more often, they might have slightly increased their scores.

The control block of trials with items at the original contrast with no noise added was run to ensure that the experi-

Table 2. Summary of Intelligibility Results (All Stimuli)

Contrast	% Correct Response for the Following Signal-to-Noise Ratio[a] 0.25	0.50	1.00	Infinite[b]
1			95.83 (72)	92.31 (104)
0.5	17.86 (112)	79.46 (112)	91.96 (112)	
0.25	27.68 (112)	80.18 (111)	86.61 (112)	
0.125	24.32 (111)	78.38 (111)	87.50 (112)	
Average Probab.	0.23	0.79	0.90	

[a] The number of trials is given in parentheses next to each value.
[b] The actual maximum s/n was ≈64. The number of trials is given in parentheses next to each value.

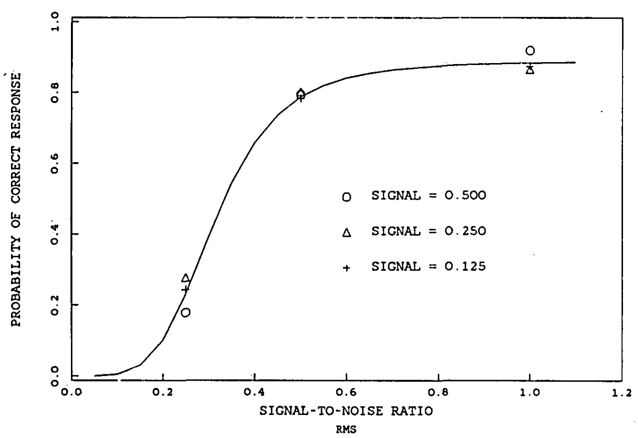

Fig. 3. Probability of correct response (averaged over subjects) as a function of the signal-to-noise ratio. Within each vertical group of points, the normalized contrast C of the images varies. The smooth curve is derived from Eq. (13).

mental procedure did not impair the intelligibility of the stimulus materials. The results (Table 2) indicate that subjects were able to perform at 92% correctness, which is slightly better than is usually observed with such tests.[13,14] In the experimental conditions, group performance ranged from a low of 18% to a high of 92%, which indicates that the experiment successfully spanned a wide range of stimulus conditions.

The Effect of the Signal-to-Noise Ratio

Table 2 and Fig. 3 indicate that for each contrast level, the probability of correct response is a monotonically increasing function of the signal-to-noise ratio. The contrast alone does not seem to have any systematic effect. The largest difference (10%) that was due to overall contrast (with s/n held constant) occurred at the lowest s/n levels, but the effects of contrast on performance were not systematic.

We tested the signal-to-noise hypothesis by using a χ^2 test. The test was based on the consequence of the signal-to-noise hypothesis that the probability of correct identification should be the same for the same signal-to-noise ratio. In the main 3 × 3 study, there were three different levels of s/n, each composed of three different combinations of signal and noise. For each s/n, we computed the average intelligibility and used that as the expected value for computing χ^2. The χ^2 for each s/n has two degrees of freedom, and the sum of all χ^2 then has six degrees of freedom. The resulting, statistically insignificant, $\chi^2 = 5.08$ is expected under the signal-to-noise hypothesis. For comparison, we computed χ^2 on the basis of other hypotheses but could not find an alternative that we would not reject. Over the range of s and n values studied, only s/n (and not s or n individually) determined performance.

Limits of Visual Communication: A Functional Formulation

Having confirmed that the signal-to-noise ratio is the most important predictor of performance, we propose an additional analysis to demonstrate how this performance is derived from the interplay of three factors: (1) s/n, (2) the subjects' capabilities, and (3) the difficulty of the ASL signs.

To begin, consider the fact that observers' performance is limited at both low and high signal levels. The low-signal limitations, which are primarily due to the observers' inability to see the images clearly, have been traditionally characterized by a constant internal noise that is added to the signal (e.g., Ref. 26). In what follows we will ignore this noise, since we have not measured intelligibility at low signal levels. The high-level limitations may be caused by a number of factors, including the observers' expertise, their cognitive abilities, the implementation of the signs, and the quantizing noise. For the sake of uniformity we developed a model in which the the high-signal limits are also represented by additive noise.

Thus the following development of a model is based on two simplifying assumptions: (1) the variances of all noises, internal and external, simply add together to produce the effective noise variance and (2) the probability of a correct identification is directly proportional to the effective signal-to-noise ratio, that is, the ratio of the signal to the effective noise. The second assumption can be elaborated with two additional parameters in order to fit the data. We assume a direct proportionality because it considerably simplifies the resulting functional form. Other assumptions could equally well have been made.

Under the first assumption (that an observer's limitations can be represented by additive sources of internal noise n_i), using the terminology of Pelli,[26] we define the effective noise n_e as the sum of the internal and external noises $n_e^2 = n^2 + n_i^2$. The effective signal-to-noise ratio s/n_e that determines performance is then the ratio of the signal to the effective noise: $s/n_e = s/(n^2 + n_i^2)^{1/2}$.

The assumption that the performance is proportional to s/n_e, together with the empirically observed upper bound on performance as s increases, determines an asymptotic limit of s/n_e. Let the upper bound on identification performance be $1/k$, $k \geq 1$. The limit

$$\lim_{s\to\infty} \frac{s}{(n^2 + n_i^2)^{1/2}} = \frac{1}{k},$$

indicates that, for large signal levels, the magnitude of the internal noise is proportional to the signal strength, $n_i = ks$. The effective noise then becomes

$$n_e^2 = n^2 + (ks)^2. \tag{5}$$

The second assumption, that the performance p is proportional to the effective signal-to-noise ratio, can be written as

$$p(s, n) = \left(\frac{s}{n_e}\right)^2 = \frac{s^2}{n^2 + (ks)^2} = \frac{1}{k^2 + \left(\frac{n}{s}\right)^2}. \tag{6}$$

Although p could, in principle, be a function of both s and n, in this formulation it is a function of only one variable, the external signal-to-noise ratio. When s/n is expressed in terms of its logarithm,

$$p\left(\frac{s}{n}\right) = \frac{1}{k + \epsilon^{-2\log(s/n)}}, \tag{7}$$

where $k \geq 1$ is a real constant, then the performance function in this equation has the basic form of the logistic function. (Note that the logistic distribution function is quite similar to the Gaussian distribution function but has a more convenient and more appropriate functional form for the present analysis.) To characterize actual data, two additional parameters are required: a parameter α for scaling and a parameter β for shifting. Introduction of these parameters into Eq. (7) yields

$$P\left(\frac{s}{n}\right) = \frac{1}{k + \epsilon^{-\alpha\log(s/n)-\beta}}, \tag{8}$$

where $\alpha > 0$, β, and $k > 1$ are real constants and P is the the probability of correct identification of an ASL sign. [Within a probabilistic interpretation of P (i.e., logistic distribution of a random variable), β and α will be closely related to the mean (expected value) μ and the variance σ, respectively.]

The origin of internal noise and its dependence on signal level deserves a discussion. We consider two factors (quantization error and ignorance) that are modeled well by signal-correlated noise.

Quantization Noise

The images in our experiment were quantized to a fixed number of levels (in particular, 16), and then the signal amplitude was amplified. This procedure was used in order to ensure that the luminance signal was identical at all signal levels. It is customary and convenient to describe such error introduced by a quantizing process as additive, quantizing noise. For a given quantized signal, the absolute amplitude of the quantizing noise, i.e., the difference between the original signal and the quantized signal, will increase with the amplification factor but will always be proportional to the signal level.

By using standard statistical assumptions, we estimated the average signal-to-noise ratio in our stimuli that was due to quantizing noise. The resulting lower bound on the rms s/n ratio is approximately 6.9 (46.9 in power units), which is well above the highest signal-to-noise ratio used in our experiment. The average effect of the quantizing noise would be negligible in the experiments.

Representing Ignorance as Noise

A second component of internal noise arises from a combination of an observer's unfamiliarity with particular ASL signs, with their implementation in our test, or with their English glosses. To represent the observer's ignorance as additive noise, its amplitude must be made proportional to the amplitude of the signal so that ignorance remains independent of the signal level.

In both the quantization and ignorance components of the internal noise, it is the combination of the assumption of noise additivity with the assumption of the invariance of the performance with the signal level that requires proportionality between the noise amplitude and the signal level. This proportionality is embodied in the constant k. A multiplicative combination rule would not require dependency of the noise on the signal level, but the resulting proportion of signal variance that is due to noise would be the same. The reason for choosing an additive representation is that, in the present experiments, we investigated the effects of additive external noise and compared the additive effects of this external noise with the internal noise.

Detection performance has been characterized traditionally by the slope and the location of the corresponding psychometric functions. The psychometric functions are, in turn, assumed to describe probability distributions of underlying random variables. To interpret P [Eq. (11)] in terms of an underlying probability distribution, we consider the conditional probability that the subject will make a correct response, given that he or she is familiar with the particular ASL stimulus; i.e., we consider k. This may be interpreted as a cumulative probability-distribution function of a random variable that determines the intelligibility of ASL stimuli. With this interpretation of P, the parameters α and β represent the standard deviation $\sigma = 1/1.814\alpha$ and the mean $\mu = \beta/\alpha$, respectively, of the underlying random variable. The actual values of μ and σ are expressed in units of $\log(s/n)$, and they do not depend on k.

The constants β, α, and k were computed by using the functional form of Eq. (8) in conjunction with a maximum-likelihood estimation technique. Note that these estimates are based on the rms values of the ratio s/n. The resulting parameter values are

$$\alpha = 4.476,$$

$$\beta = 5.033,$$

corresponding to the following distributional parameters:

$$\mu = -1.124,$$

$$\sigma = 0.123,$$

$$k = 1.125.$$

The psychometric function that is based on these parameter values was plotted in Fig. 3. It shows that $\mu = -1.124$ is the value of $\log(s/n)$ at which 0.5 of the asymptotic performance is reached, that the steepest slope is σ, and that the asymptotic performance level is k.

The convenience of this μ, σ, k representation justifies the algebraic development; our data are not reduced by the logistic representation. However, the logistic representation has another important advantage. Because the main goal of this experiment was to determine the invariance of performance with a constant signal-to-noise ratio even as the stimulus contrast varied, it was not practical to test empirically each ASL sign in each condition. Therefore determining the difficulty of individual ASL signs requires a functional model to relate signs that were never tested in the same condition. For example, 29 of the 87 distinct ASL signs (one sign was repeated) were either always or never recognized. Did recognition failures occur because ASL signs were difficult or because they occurred only in difficult conditions? Because conditions differ only in objectively measurable additive noise, the parameters of the logistic representation can be estimated and used to estimate the difficulty of individual signs and conditions (noise) even when no signs occur in more than one condition; that is, the logistic representation permits the segregation of the item difficulty and the condition difficulty in a seemingly intractable case. (See the Discussion section for full details.)

Effects of Item and Subject Differences: Subject Differences

A number of *post hoc* analyses were performed to examine other possible sources of variability. These analyses sought to determine the source of differences in difficulty between stimuli and to estimate the effects of differences in skills between the individual observers.

Individual Differences

An analysis was performed to determine whether there was evidence in our performance data for an interaction between subjects and the effects of the signal-to-noise ratio. Each subject's performance was analyzed separately, and the results were tested for interaction by using a two-way analysis of variance. There was no significant interaction, and we could not find any systematic effect for the individual subjects. All our subjects were affected quantitatively in the same way by stimulus degradation.

Word Frequency

The performance in auditory intelligibility experiments with unspecified word sets depends on the prior frequency of the words in natural language,[30] high-frequency words being identified at lower s/n levels. By analogy, we would expect that a sign with a high frequency of usage in ASL may be identified at lower s/n than an uncommon sign. To examine this sign-frequency hypothesis, it would be necessary to have norms for the frequency of usage of signs. Such information is at present unavailable. Therefore we assumed that the frequency of signs follows patterns similar to that of English. We compared the performance on each sign with the frequency of the nearest corresponding English word. The scatter diagram of this comparison is shown in Fig. 4. The correlation between the frequency and overall performance is small (0.153) and not statistically significant. It may be that the correlations are small in part because the signs in the study all were relatively high-frequency signs. Confirmation of this result awaits the availability of frequency norms for ASL signs.

Removing Perfect Stimuli

A closer examination of the data revealed that 24 of the 88 stimuli were correctly identified on every presentation. To ensure that the signal-to-noise-ratio results were not obtained by the particular mixture of difficult and easy stimuli, we repeated all the data analyses with those perfect stimuli excluded. The results and the conclusion remained unchanged: the intelligibility score was determined by the signal-to-noise ratio.

Effects of Item and Subject Differences: Item Differences

Components of Stimulus Difficulty

Up to the preceding paragraph, we had assumed, in our analyses, that all signs were about equally easy or difficult to identify. Even a superficial examination of the data, however, suggests considerable differences among the signs. Some signs were recognized quite accurately, whereas others were not. The determination of the actual difficulty of each individual sign was not possible using the raw data because the different ASL signs were presented with different s/n ratios. Here, we take advantage of the general form of the functional relationship between the probability of correct identification and s/n to evaluate the difficulties of individual stimuli.

There are at least two potential components of difficulty in identifying individual signs. One is observers' possible lack of familiarity with some of the ASL signs. This effect cannot be eliminated by increasing the signal level or the signal-to-noise ratio. This component was discussed earlier and was summarized by the constant k in Eq. (8). The second component, a perceptual one, arises from the limits of visual capabilities and information content of the images. The purpose of the following analysis is to describe the difficulty of individual signs in terms of signal related parameters.

Representation of the Difficulty of ASL Signs

Since the signal-to-noise ratio was the predominant determinant of intelligibility, our hypothesis was that the perceptual component of difficulty of an ASL sign could be expressed as a parameter modifying the value of the physical s/n. There are several ways to define such a parameter (e.g., additive to s/n, additive to s or n), and we explored several reasonable formulations of item difficulty. The best-fitting representation of sign difficulty consisted of a constant $d(x)$ representing the sign difficulty that was subtracted from the logarithm of s/n; alternatively, $d(x)$ may be viewed as an efficiency factor that multiplies s, with the most difficult signs being inefficient and requiring more s.

The reader may recall that the constant k in Eq. (8) represented both the sign familiarity and the effects of quantizing noise over the ensemble of signs. In order to characterize

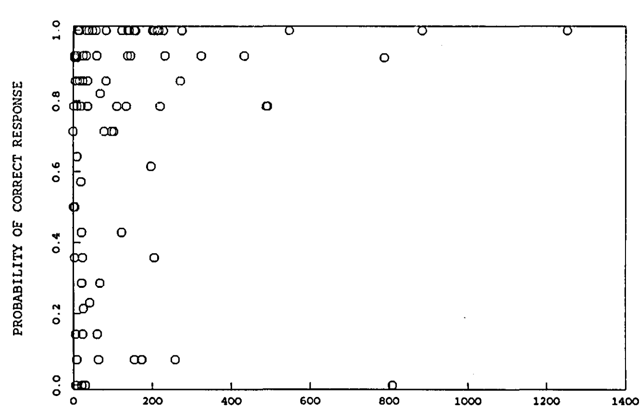

Fig. 4. Probability of a correct response (intelligibility) as a function of the English word frequency for each of the ASL signs tested. For these data, the correlation between intelligibility and word frequency is 0.15, which is not statistically different from zero.

signs whose probability of correct identification exceeded $1/k$, we computed the individual sign difficulty using Eq. (8) with $k = 1$. The intelligibility of an ASL sign x then depends on s/n and on the difficulty parameter $\mathrm{d}(x)$ as follows:

$$P\left(x, \frac{s}{n}\right) = \frac{1}{1 + \epsilon^{-\alpha[\log(s/n) - \mathrm{d}(x)] - \beta}}. \tag{9}$$

The resulting values $\mathrm{d}(x)$ (i.e., logarithm of the losses) are shown in Table 1. The fit of this model was assessed on those signs whose probability of identification was less than one and greater than zero. Each value of $\mathrm{d}(x)$ was used to predict the difficulty of the same sign in up to three different s/n conditions. The model was tested by using a χ^2 test; the resulting $\chi^2 = 24.2$ had 21 degrees of freedom. Note that for the stimuli that were either always or never correctly identified (100% or 0% correct), we can determine only the upper and lower bounds of $\mathrm{d}(x)$, respectively. These bounds are identified by the corresponding < and > symbols in Table 1. The reason that these bounds may differ for two stimuli with the same probability of correct response is that they were used with different s/n levels. For example, consider two signs that were always correctly recognized; the sign shown in the low s/n condition will have a lower difficulty bound than the one that was shown with a high s/n. Similarly, the difficulty of the signs does not necessarily increase with the probability of error. There are signs (e.g., "accident") that were identified with high probability (93%) whose difficulty (0.552) is higher than other signs (e.g., "behind") that were almost never identified (7%) but whose difficulty is estimated to be lower (0.311). The reason for this nonmonotonicity is that the former sign was presented in conditions with a higher s/n ratio than the latter. It is worth noting that this type of inference would not be possible without the model.

The important result of this analysis is that the effect of the perceptual component of difficulty can be viewed as a reduction of the signal-to-noise ratio.

The Relation between s/n and Information Capacity

The most important result of this study is the confirmation of the signal-to-noise hypothesis. However, in the process, we measured the quantitative relationship between ASL intelligibility and the signal-to-noise ratio. Shannon[22] proved that, under certain assumptions, s/n and the information capacity c of a channel are related by $c = TW \log_2[(s^2 + n^2)/n^2]$, where W is the spatial bandwidth of the signal and the noise and where T is time. It is tempting to combine the measured performance at various values of s/n with Shannon's theorem to infer the noise of the human perceptual channel and ultimately, thereby, its information capacity.

In the light of the apparent simplicity and directness of Shannon's theorem, it is instructive to confront the many obstacles between the data and estimates of perceptual information capacity. The critical value of s/n, at which performance begins to suffer as external noise is added, is often used to estimate the internal noise (e.g., Ref. 26). However, our data show that internal noise is negligible under the conditions of the experiment; otherwise, the ratio s/n would not have been adequate to describe the data; for constant s/n, performance would have suffered as s and n decreased. What we did observe is that, as s/n is increased, a critical point is reached approximately when rms $s/n = 0.5$. At this point the internal perceptual noise equals the external noise. However, this internal noise is the signal-correlated noise that is described by the parameter k of Eqs. (5) and (6) and not the signal-independent noise implied by Shannon's theorem.

Quite apart from its relation to information theory, the numerical value $s/n = 0.5$, as an estimate of equivalent internal and external noise, must be interpreted with great caution because the signal and the noise have different spectral power densities. In our study, s and n represented the total power in the signal and the noise, respectively. When the masking effect of noise is band specific, that is, when noise masks signals best if both s and n are in the same spatial-frequency band, then s/n has meaning only if both s and n have the same spectrum. For example, at the given level of s, we might have achieved the same performance reduction with a much less powerful n if it had been spectrally better placed. The spectrally optimized noise would have yielded a lower estimate of internal noise (higher critical s/n) and hence a correspondingly higher estimate of human perceptual capacity. To obtain a meaningful numerical value for the critical s/n, it is minimally necessary to ensure that s and n have the same spatiotemporal-frequency spectrum. Ultimately, it is better to divide the spatial–temporal frequency spectrum into relatively narrow bands and to investigate performance band by band rather than to coalesce the frequency bands in both the signal and the noise. Indeed, this is the approach taken by Pelli[31] and Riedl[32] and is the arduous but only available path toward theoretically relating s/n to perceptual information-processing capacity.

SUMMARY AND CONCLUSIONS

(1) An intelligibility test was developed for video communication of ASL. The test consists of carefully selected ASL signs and a procedure for determining observers' responses.

(2) When noise was added to ASL images, the intelligibility of ASL was predominantly determined by the signal-to-noise ratio, independently of the signal and the noise levels within the range tested.

(3) A logisticlike function was proposed to characterize the relationship between intelligibility and the signal-to-noise ratio. The logistic function was chosen on the basis of algebraic simplicity; other similar distributions might have served equally well.

(4) The difficulty of individual ASL signs can be represented as consisting of two additive noise components: one, which is due to sign familiarity and quantizing noise, behaves as signal-correlated noise (added noise that is proportional to the signal) and implies an upper limit in intelligibility; the other, a perceptual-difficulty component, behaves as noise added to $\log(s/n)$; it acts as an attenuation of the physical signal and hence of s/n.

ACKNOWLEDGMENTS

The experimental work for this project was carried out at the Human Information Processing Laboratory in the Department of Psychology of New York University and supported by the National Science Foundation, Science and Technology to Aid the Handicapped grant PFR-80171189. Subsequent analyses of the data and the preparation of the final report were carried out jointly at New York University and at Stanford University, supported in part by U.S. Air Force Life Sciences Directorate grants AFOSR 80-0279 and 85-0364. We appreciate and gratefully acknowledge the assistance that we received from many persons in the deaf community and schools for the deaf, including Jerome Schein, formerly director of the Deafness Research and Training Center at New York University; the staff of Public School 47; the New York Society for the Deaf; and the Hebrew Association of the Deaf. Nancy Frishberg provided essential guidance in the construction of the intelligibility test materials. We thank our patient deaf signer, Ellen Roth, and acknowledge the many useful comments of Michael S. Landy and the skillful technical assistance of Robert Picardi.

REFERENCES

1. G. E. Legge, G. S. Rubin, D. G. Pelli, and M. M. Schleske, "Psychophysics of reading. II. Low vision," Vision Res. **25**, 253–266 (1985).
2. D. G. Pelli, Institute for Sensory Research, Syracuse University, Syracuse, New York 13244-5290 (personal communication, 1983).
3. D. G. Pelli and L. C. Applegate, "The visual requirements of mobility," Invest. Ophthal. Vis. Sci. Suppl. **26**, 57 (1985).
4. D. Regan and D. Neima, "Low-contrast letter charts as a test of visual function," Ophthalmology **90**, 1192–1200 (1983).
5. A. P. Ginsburg, "Visual information processing based on spatial filters constrained by biological data," Vols. 1 and 2 of doctoral dissertation (Cambridge University, Cambridge, 1978); also in Lab. Rep. AMRL-TR-78-129 (Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio, 1978).
6. A. P. Ginsburg, "Specifying relevant spatial information for image evaluation and display designs: an explanation of how we see certain objects," Proc. Soc. Inf. Disp. **21**, 219–227 (1980).
7. G. E. Legge, G. S. Rubin, D. G. Pelli, and M. M. Schleske, "Psychophysics of reading. I. Normal vision," Vision Res. **25**, 239–252 (1985).
8. G. S. Rubin and K. Siegel, "Recognition of low-pass filtered faces and letters," Invest. Ophthalmol. Vis. Sci. Suppl. **25**, 71 (1984).
9. J. F. Abramatic, P. Letellier, and M. Nadler, "A narrow-band video communication system for the transmission of sign language over ordinary telephone lines," in *Image Sequence Processing and Dynamic Scene Analysis*, T. S. Huang, ed. (Springer-Verlag, Berlin, 1983), pp. 314–336.
10. D. E. Pearson, "Visual communication systems for the deaf," IEEE Trans. Commun. **C-29**, 1986–1992 (1981).
11. D. E. Pearson, "Evaluation of feature-extracted images for deaf communication," Electron. Lett. **19**, 629–631 (1983).
12. D. E. Pearson and H. Six, "Low data-rate moving-image transmission for deaf communication, in *Proceedings of the International Conference on Electronic Image Processing* (Institution of Electrical Engineers, London, 1982), pp. 204–208.
13. G. Sperling, "Bandwidth requirements for video transmission of American sign language and finger spelling," Science **210**, 797–799 (1980).
14. G. Sperling, "Video transmission of American sign language and finger spelling: present and projected bandwidth requirements," IEEE Trans. Commun. **C-29**, 1993–2002 (1981).
15. G. Sperling, M. Pavel, Y. Cohen, M. S. Landy, and B. J. Schwartz, "Image processing in perception and cognition," in *Physical and Biological Processing of Images. Rank Prize Funds International Symposium at The Royal Society of London*, O. J. Braddick and A. C. Sleigh, eds., Vol. 11 of Springer Series in Information Sciences (Springer-Verlag, Berlin, 1983), pp. 359–378.
16. G. Sperling, M. S. Landy, Y. Cohen, and M. Pavel, "Intelligible encoding of ASL image sequences at extremely low information rates," Comput. Vision Graphics Image Process. **31**, 335–391 (1985).
17. A. E. Burgess, "Visual signal detection. III. On Bayesian use of prior knowledge and cross correlation," J. Opt. Soc. Am. A **2**, 1498–1507 (1985).
18. A. E. Burgess, R. F. Wagner, R. J. Jennings, and H. B. Barlow, "Efficiency of human visual detection," Science **214**, 93–94 (1981).
19. U. Bellugi and S. Fischer, "A comparison of sign language and spoken language," Cognition **1**, 173–200 (1972).
20. J. E. Hawkins and S. S. Stevens, "The masking of pure tones and of speech by white noise," J. Acoust. Soc. Am. **22**, 6–13 (1950).
21. H. Poizner and H. Lane, "Discrimination of location in American sign language," in *Understanding Language through Sign Language*, P. Siple, ed. (Academic, New York, 1978), pp. 271–287.
22. C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (U. of Illinois Press, Urbana, Ill., 1949).
23. D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966).
24. E. H. Weber, *De Pulsu, Resorptione, Auditu, et Tactu* (Koehler, Leipzig, 1834).
25. R. S. Woodworth and H. Schlosberg, *Experimental Psychology* (Holt, New York, 1954).
26. D. G. Pelli, "Effects of visual noise," doctoral dissertation (Cambridge University, Cambridge, 1980).
27. M. S. Landy, Y. Cohen, and G. Sperling, "HIPS: a UNIX-based image processing system," Comput. Vision Graphics Image Process. **25**, 331–347 (1984). (HIPS is the Human Information Processing Laboratory's image-processing system.)
28. M. S. Landy, Y. Cohen, and G. Sperling, "HIPS: Image processing under UNIX software and applications," Behav. Res. Methods Instrum. **16**, 199–216 (1984).
29. Y. Cohen, "Measurement of video noise," Tech. Rep. (Human Information Processing Laboratory, New York University, New York, 1982).
30. I. Pollack, "Message uncertainty and message reception," J. Acoust. Soc. Am. **31**, 1500–1508 (1959).
31. D. G. Pelli, "The spatiotemporal spectrum of the equivalent noise of human vision," Invest. Ophthalmol. Vis. Sci. Suppl. **24**, 46 (1983).
32. T. R. Riedl, "Spatial frequency selectivity and higher level human information processing," doctoral dissertation (Department of Psychology, New York University, New York, 1985).